Differences in semantics explain differences in comprehension: Generic masculines and gender star forms in German from a discriminative perspective

Dominic Schmitz

August 16, 2024

Abstract

The public debate on gender-inclusive forms in German mainly revolves around two distinct forms. The generic masculine, i.e. grammatically masculine forms used to refer to referents irrespective of their gender, and gender star forms, i.e. forms that are constructed using a novel suffix including an asterisk. While research has repeatedly shown that generic masculines apparently come with a male bias, only little research is available on gender star forms. The present paper provides a first account of gender star form semantics and relates these semantics to those of generic masculines as well as to those of specific masculines and feminines. Further, measures extracted from implementations of linear discriminative learning are used to provide detailed insights into the comprehension process and to re-analyse data from three previous studies on generic masculines and gender star forms. It is shown that generic masculines seem to be biased towards male readings, while gender star forms seem to be somewhat biased towards female readings, and that measures on their semantics and comprehension can account for differences in reaction times for continuation judgements, proportions of male and female exemplars in exemplar recall, and ratios of men vs. women in social and occupational groups. Overall, gender star forms appear to be more gender-inclusive than generic masculines.

Keywords: gender bias, generic masculine, gender star, gender-inclusive language, discriminative learning

1 Introduction

There is an ongoing debate on whether German as a language is gender-inclusive and, if not, which measures can be taken to reach linguistic gender-inclusiveness. At the centre of this debate, one finds the so-called *generisches Maskulinum* ‘generic masculine’. This term refers to grammatically masculine forms that are used to refer to individuals irrespective of their gender. Importantly, this only pertains to masculine forms which form minimal pairs with their feminine counterparts, not to epicene forms (cf. Diewald 2018). Take, for instance, the word *Lehrer* 'teacher'. In its masculine form, *Lehrer* may be used to refer to a male teacher, just as it may be used to refer to a teacher of any gender. Its feminine counterpart, *Lehrerin*, is used to refer to female teachers only.

Research on the semantic and psycholinguistic reality of such generic masculines has repeatedly found evidence for a male bias in such forms (e.g. Braun et al. 1998; Stahlberg & Sczesny 2001; Gygax et al. 2008; Misersky, Majid & Snijders 2019; Keith, Hartwig & Richter 2022; Schmitz 2024). That is, even though a masculine may be used generically and with gender-inclusive intentions, its associations and language comprehension will be skewed towards male referents. While such findings do not take away from the surface intention of the generic masculine to be a gender-inclusive form – an intention that can be traced through various stages of the German language (cf. Trutkowski & Weiß 2023) – they do call into question the gender-inclusiveness of generic masculines and with that motivate language users to invent novel and supposedly truly gender-inclusive forms.

One group of such novel forms is formed by the use of special characters in new suffixes. Combining either an underscore, colon, or asterisk with the feminine suffix *-in*, novel suffixes are formed: *-_in*, *-:in*, and *-*in* (cf. Völkening 2022). The latter is commonly known as *Gendersternchen* 'gender star (diminutive)', and will henceforth be referred to as 'gender star form'. The intended meaning of all three suffixes is the inclusion of all genders, including those beyond the binary. As an example, take *Lehrer*: Its underscore, colon, and gender star forms are *Lehrer_in*, *Lehrer:in*, and *Lehrer*in* 'teacher (of any gender)', respectively.

As these latter forms developed rather recently, there is only little research available on their linguistic features and, thus far, there is no research on their semantics available. The aim of the present study is to add to the body of research on such forms and to provide a first account of their semantics, as well as to allow further insight into the comprehension of such forms by use of computational methods.

The remainder of this paper is structured as follows. Section 2 will provide detailed descriptions of generic masculines and the novel suffix forms in German, and introduce the research questions the present paper aims to answer. Sections 3 to 5 will present the individual studies that make up this paper, including their pertinent methodological aspects, analyses, and results. In Section 3, a distributional semantic account of the gender star form, the generic masculine, and the specific masculine and feminine will be given. In Section 4, the comprehension of gender star forms will be modelled using a discriminative learning approach. In Section 5, the results of the discriminative learning approach will be used to reanalyse the linguistic data of previous studies, allowing a more in-depth account of their findings. A discussion bringing together the individual studies will be given in Section 6 and Section 7 will conclude this paper.

2 Gender-(non)inclusive language in German

2.1 The generic masculine in German

German is a grammatical gender language (Gygax et al. 2019; Stahlberg et al. 2007). That is, all nouns of the language are assigned a grammatical gender. For instance, *die Sonne* 'the sun' is grammatically feminine, *der Mond* 'the moon' is grammatically masculine, and *das Firmament* 'the firmament' is grammatically neuter. Based on the grammatical gender of the noun, the appropriate forms of dependent words, i.e. of articles, attributive adjectives, ordinal numbers, participles, adjectival, relative and question pronouns, and third-person singular pronouns (Jarnatowskaja 1968), are selected. However, grammatical gender is but one of three types of gender encoded in German nouns. A second source is lexico-semantic gender. Lexico-semantic gender describes the notion that the gender of a referent is part of the word's semantics. For example, *die Tante* 'the aunt' refers to a female sibling of a parent, while *der Onkel* 'the uncle' refers to a male sibling of a parent. A third source is conceptual gender, i.e. at the conceptual level, words are associated with stereotypes. Take, as an example, the role nouns *der Programmierer* 'the programmer' and *der Florist* 'the florist'. Both words are grammatically masculine, but differ in regard to their stereotypicality: programmers are stereotypically associated to be male, while florists are stereotypically associated to be female (cf. Misersky et al. 2014).

While the assignment of grammatical gender is mostly arbitrary for inanimate nouns, it is not at all arbitrary for animate nouns, i.e. nouns with lexico-semantic gender information. As shown above, aunts are grammatically feminine, uncles are grammatically masculine. Similarly, most German role nouns show a masculine and a feminine form. For instance, *Lehrer* 'teacher.MASC' and *Lehrerin* 'teacher.FEM'. Straightforwardly, one may correctly assume that the masculine form is used to refer to male teachers, whereas the feminine form is used to refer to female teachers. This specific use of role nouns is well attested. However, for masculine role nouns, a further use, a generic one, is found. That is, grammatically masculine role nouns like *Lehrer* are used to refer to referents of any gender or irrespective of their gender. A grammatically masculine role noun which is used in such a generic way is commonly referred to as a *generisches Maskulinum* 'generic masculine' (Diewald 2018).

Traditionally, it is assumed that grammatically masculine role nouns used as generic masculines are gender-inclusive (cf. Doleschal 2002). However, the usage of generic masculines is not without a long history of debate about its appropriateness (e.g. Hanitzsch 2021). Already in the 1970s, linguists argued that generic masculines lead to a linguistic under-representation of women (Pusch 1984; Trömel-Plötz 1982). Since then, a multitude of empirical studies brought forward evidence in favour of this claim: generic masculines were shown to induce a male bias instead of balanced mental representations of genders (e.g. Braun et al. 1998; Rothermund 1998; Rothmund & Scheele 2004; Stahlberg & Sczesny 2001; Stahlberg, Sczesny & Braun 2001; Gygax et al. 2008; Irmen & Kurovskaja 2010; Misersky, Majid & Snijders 2019; Keith, Hartwig & Richter 2022; Schunack & Binanzer 2022; Schmitz, Schneider & Esser 2023; Schmitz 2024).

2.2 Gender-inclusive forms using novel suffixes

The male bias of generic masculines has lead to the invention of a number of alternative strategies towards a more gender-inclusive German language. For strategies from feminisation such as the pair form, e.g. *Lehrerin und Lehrer* '(female) teacher and (male) teacher', to neutralisation such as nominalised forms, e.g. *Lehrkraft*, to the invention of novel gender suffixes, e.g. *Lehrer_in*, *Lehrer:in*, and *Lehrer*in* 'teacher (of any gender)', it was found that such strategies decrease the male bias reported for generic masculines (e.g. Gabriel et al. 2008; Schunack & Binanzer 2022).

However, regarding the level of gender-inclusiveness, such strategies differ. Pair forms consist of the grammatical masculine and feminine forms of a word, representing male and female referents. Similarly, the shorthand version of pair forms, i.e. the combination of hyphen and slash as in *Lehrer-/in*, is restricted to binary gender representation. Strategies that go beyond the representation of a gender-binary are nominalised forms as well as the forms using novel suffixes. Nominalised forms mostly make use of either compounding, e.g. *Lehrkraft* 'teacher', or participles, e.g. *Lehrende* lit. 'those who are teaching'. The downside of such forms are that compounds do not exist for all role nouns, that participles generally, i.e. before a potential lexicalisation, carry the semantics of progressiveness, and that participles only work gender-inclusively in the plural, as in the singular the words that are dependent on the participles' grammatical gender will require either the masculine or feminine for form selection.

Forms that make use of novel suffixes with underscores, colons, or asterisks are also intended to include genders beyond the binary. The morphology of such forms is rather straightforward. While the grammatically feminine form of a role noun is derived by adding the feminine suffix *-in* to a grammatically masculine role noun, these novel forms are analogously derived by adding the respective novel suffix, i.e. *-_in*, *-:in*, or *-*in*, to a grammatically masculine role noun (cf. Völkening 2022). Take, again, as an example *Lehrer* 'teacher'. Its feminine counterpart is *Lehrerin*, its counterparts with the novel suffixes are *Lehrer_in*, *Lehrer:in*, and *Lehrer*in*. While the morphological process for such role noun forms themselves is rather simple, morphosyntactical issues may arise in certain contexts. That is, somewhat similar to the issue of participles as gender-inclusive forms, words that are dependent on the role noun's grammatical gender will require the selection of their appropriate form. In some cases, e.g. for articles, this may be less straightforward than for the role nouns themselves (Völkening 2022).

Thus far, only a few studies investigated the linguistic features of the forms using the novel suffixes, focussing mostly on either their morphological and morphosyntactic features (e.g. Völkening 2022) or on their comprehension (e.g. Zacharski & Ferstl 2023; Körner et al. 2022; Kurz & Mulder 2023; Schunack & Binanzer 2022).[1] Zacharski & Ferstl (2023) investigated the impact of the gender star form on mental representations using a word-picture matching task. Participants assessed the suitability of images depicting male, female, and non-binary individuals as illustrations of role nouns presented in masculine, feminine, or gender star form. The findings show that masculine nouns were not automatically read as gender-neutral, evidenced by processing difficulties for female images following masculine nouns. Conversely, high acceptance rates and faster reaction times for all images following the star form indicate that it effectively activates inclusive mental representations encompassing male, female, and non-binary individuals. This suggests that the gender star elicits the intended inclusive interpretation, supporting its use for increasing visibility of non-binary persons in the German language.

Körner et al. (2022) conducted two judgement tasks on sentence continuations. Between participants, the target word form in the first sentence varied between generic masculines, gender star forms, pair forms with the masculine form first, and pair forms with the feminine form first. All four forms were used is the first task, while the second task only made use of generic masculines and gender star forms. The initial sentences were followed by a second sentence, which within participants varied between male and female referents. The participants' task was to evaluate whether the second sentence was a meaningful continuation of the first. The authors analysed continuation judgements, i.e. the share of yes/no answers, and the reaction times of these answers. They found that following masculine generics, continuations with male referents were judged more frequently and faster to be meaningful continuations, while after gender star forms continuations with female referents were judged more frequently (both tasks) and faster (second task) as meaningful continuations. That is, the findings are in line with the male bias for generic masculines found in previous studies and indicate a female bias for the gender star form.

Kurz & Mulder (2023) conducted a study following the general design by Stahlberg & Sczesny (2001). In an online questionnaire, participants were asked to provide three exemplars of six categories: politicians, athletes, singers, TV hosts, actors, and authors. In their between-subjects-design, categories were either presented as generic masculines, e.g. *Politiker* 'politicians', or as gender star forms, e.g. *Politiker*innen* 'politicians'. Kurz & Mulder (2023) found that participants who encountered the categories represented by gender star forms provided more female exemplars than those who encountered generic masculines. These findings are in line with the male bias found for

[1] Note that there are even fewer studies on the spoken version of such novel forms, like the one by Körner, Glim & Rummer (2024).

generic masculines in previous studies.

Schunack & Binanzer (2022) replicated earlier studies from the field of gender-inclusive language by Stahlberg & Sczesny (2001) and Gabriel et al. (2008). While their adapted replication of the study by Stahlberg & Sczesny (2001) did not include the gender star form, one of the two adapted replications of the experiment by Gabriel et al. (2008) did. In an online questionnaire, participants were asked to indicate the ratio of men vs. women in the social and occupational groups represented by the target word paradigms, based on their personal assumptions. Paradigms were chosen following results by Misersky et al. (2014) on stereotypicality, grouping them into stereotypically female, male, and neutral categories, and presented between participants either as generic masculines, capital I forms,[2] or gender star forms. Schunack & Binanzer (2022) found that with capital I forms and gender star forms, paradigms were judged with higher female ratios than with generic masculines if they were either of female or neutral stereotypicality. For paradigms with a male stereotypicality, only gender star forms showed an increase in female ratios. That is, the gender star form increased the ratio of assumed women in a given group significantly independently of the group's stereotypicality.

2.3 Research questions

While the body of previous research on forms with novel suffixes already shed some light on the comprehension of such forms, they only offer an account of such forms direct impact on elicited measures, e.g. ratios of word-picture matching judgements, reaction times for continuation judgements, proportions of male and female exemplars in exemplar recall, and ratios of men vs. women in social and occupational groups. What is lacking is an account of how the effects of such forms on the elicited measures come to be. To fill this gap, the present paper aims at answering three questions that build on each other.

First, a general account of the semantics of generic masculines and gender star forms in comparison to specific masculines and feminines is needed. This account will be given using methods of distributional semantics (e.g. Harris 1954: see Sections 3.1 and 3.3). With different semantics, forms should naturally show differences in elicited measures:

> *RQ1: How similar or dissimilar are the semantics of specific masculines, specific feminines, generic masculines, and gender star forms?*

Second, in order to gain further insight into the semantics of generic masculines and gender star forms, their semantics are analysed making use of linear discriminative learning (Baayen et al. 2019), which allows the computation of different semantic measures (see Sections 4.1 and 4.3).

[2] Capital I forms are constructed using the feminine suffix -*in* but with a capitalised *I*. They are meant to represent both binary genders.

RQ2: How do the semantics of generic masculines and gender star forms differ in detail?

Third, the semantic measures are directly related to the findings of previous studies (Körner et al. 2022; Kurz & Mulder 2023; Schunack & Binanzer 2022) to uncover how the fine semantic differences of generic masculines and gender star forms impact the measures elicited in these studies (see Section 4.3).

RQ3: Which semantic measures account for the differences found between generic masculines and gender star forms in previous studies?

3 Study 1: A distributional semantic account of gender star forms

3.1 Corpus creation, target paradigms, and semantic vectors

For the computation of semantic vectors of generic masculines, specific masculines, specific feminines, and gender star forms, by means of distributional semantics (cf. Boleda 2020), a corpus containing attestations of such forms is required. As there is no such corpus readily available, a new corpus was built using pre-existent as well as novel parts.

The first part of the corpus consisted of the corpus by Schmitz, Schneider & Esser (2023). This corpus contained 30,000 sentences with attestations of generic masculines, specific masculines, and specific feminines. Importantly, all attestations were readily annotated for genericity, i.e. whether a masculine role noun was used generically or specifically. As there is no automated annotation of genericity in German role nouns available, it was decided to use the readily annotated attestations of Schmitz, Schneider & Esser (2023) and, with that, the 113 target paradigms of their study. Each target paradigm consisted of the generic masculine, specific masculine, and specific feminine forms of a given word in all cases and across number. An example is given in Table 1. The corpus by Schmitz, Schneider & Esser (2023) contained 3689 duplicated sentences; these were removed for the current study, resulting in 26,311 retained sentences.

The second part of the corpus consisted of online articles from *Tagesspiegel*, a German newspaper based and mostly read in Berlin. The newspaper used gender star and gender colon forms in many articles from 2020 to 2023, but stopped using such gender-inclusive forms in November 2023 due to requests by readers ("‚Tagesspiegel' schafft Gendersternchen ab" 2023). To find articles with the gender star equivalents of the 113 target nouns, Google was used with the search string 'site:tagesspiegel.de target*innen' and 'site:tagesspiegel.de target:innen'.[3] Articles with the respective gender-inclusive

[3] As explained in Section 2.2, the gender star form and the colon form can be used interchangeably. To increase the number of attestations, all colon forms were collected and colons changed into asterisks afterwards. Underscore forms were not attested.

Table 1: Paradigm of *Schüler* 'pupil'.

	singular			
	generic masculine	specific masculine	specific feminine	gender star
nominative	Schüler	Schüler	Schülerin	Schüler*in
genitive	Schülers	Schülers	Schülerin	Schüler*in
dative	Schüler	Schüler	Schülerin	Schüler*in
accusative	Schüler	Schüler	Schülerin	Schüler*in
	plural			
	generic masculine	specific masculine	specific feminine	gender star
nominative	Schüler	Schüler	Schülerinnen	Schüler*innen
genitive	Schüler	Schüler	Schülerinnen	Schüler*innen
dative	Schülern	Schülern	Schülerinnen	Schüler*innen
accusative	Schüler	Schüler	Schülerinnen	Schüler*innen

forms were retained completely, i.e. not only the sentences containing the attestations, to increase the diversity of grammatical structures and lexical content. While such articles may well contain attestations of masculine and feminine role nouns, one may assume that they are of specific nature, as the gender star and colon forms were used in place of generic masculines. As is shown in Table 1, the gender star forms constitute a further dimension for target paradigms.

The third part of the corpus consisted of random attestations from the Leipzig Corpora Collection's *news* sub-corpus (Goldhahn, Eckart & Quasthoff 2012). This is not only the source of the sentences used by and adopted from Schmitz, Schneider & Esser (2023), but also consists of snippets from news websites and is hence similar in genre to the *Tagesspiegel* language material. To further increase the amount of language material, 1,400,000 sentences were randomly sampled but controlled to not contain any attestations of target words. Samples were taken in packages of 100,000 sentences for the years 2010 to 2023. This allowed a wide variation of topics and included the years covered by both the sentences from Schmitz, Schneider & Esser (2023) and the *Tagesspiegel*.

The set of target paradigms originally corresponded to that used in Schmitz, Schneider & Esser (2023). However, four of these forms are morphologically different from the majority: they do not use the feminine suffix -*in* to form a feminine form but drop the final *r*. Analogously, gender star forms for these words are not attested. Therefore, the following targets (given in their masculine form) were disregarded: *Bankangestellter* 'bank employee', *Gefangener* 'prisoner', *Obdachloser* 'homeless person', and *Stellensuchender* 'person on a jobhunt'.

The remaining 109 target paradigms consisted of singular and plural versions of the generic masculine, specific masculine, specific feminine, and gender star forms. Across all forms, it was found that more attestations in the plural than in the singular are part of the corpus; this was especially true for gender star forms (94 % plural forms). As previous computational studies on gendered role nouns in German (Schmitz, Schneider & Esser 2023; Schmitz 2023, 2024) found that semantic similarities computed via semantic vectors are of identical nature across number, it was decided to change all attestations within the corpus to their plural forms in the nominative. That is, semantic vectors were only computed for the plural versions of the generic masculine, specific masculine, specific feminine, and gender star forms. This decision allowed to retain as much language material as possible.

The final corpus consisted of 1,624,934 sentences with 25,528,231 tokens of 55,063 types, including 11,829 generic masculine, 10,302 specific masculine, 4,180 specific feminine, and 9,093 gender star attestations of the target word paradigms. All sentences were tagged using the RNNTagger software (Schmid 1999), rendering the surroundings of masculine and feminine forms grammatically similar.

Based on the corpus, semantic vectors were computed with a *fastText* model trained using *Gensim* (Řehůřek & Sojka 2010) in Python (Python Software Foundation 2023). Such vectors are meant to capture the semantics of the word forms they belong to, following the ideas of the distributional hypothesis, which states that differences in meaning are represented by differences in distribution (Harris 1954; Boleda 2020). Thus, if words occur in different contexts, their semantics are expected to be different. If words frequently occur in similar contexts, their semantics are expected to be similar. There are several methods with different algorithms at work to arrive at semantic vectors, e.g. *Word2Vec* (Mikolov et al. 2013), *GloVe* (Pennington, Socher & Manning 2014), *NDL* (Baayen et al. 2011), or, as in the present case, *fastText* (Bojanowski et al. (2016); for a brief introduction see Schmitz (2024)).

3.2 Analysis

The similarity of the semantics of target word forms was analysed using cosine similarity as a measure. Cosine similarities were computed within target word paradigms, not across. For instance, the semantic vector of the generic masculine *Schüler* was compared to the semantic vector of its specific feminine counterpart *Schülerin*, but not to members of other target paradigm like *Sängerin* '(female) singer'.

To compare cosine similarities between target word forms efficiently, beta regression was introduced in generalised additive mixed models using the *mgcv* package (Wood 2017). Beta regression was chosen as the model of choice, because cosine similarity values were in the range of $(0, 1)$, i.e. the range of values that underlies beta regression, the beta distribution. Using beta regression in generalised additive mixed models allowed the inclusion of the following variables:

COMPARISON. This is the variable of interest, as it codes which paradigm member types are compared to each other. It takes the values `SF vs. SM` (specific feminine vs. specific masculine), `SF vs. GM` (specific feminine vs. generic masculine), `SF vs. ST` (specific feminine vs. star form), `SM vs. GM` (specific masculine vs. generic masculine), `SM vs. ST` (specific masculine vs. star form), and `GM vs. ST` (generic masculine vs. star form).

STEREOTYPICALITY. This variable contains information on the stereotypicality of a given paradigm, i.e. its stereotypical load regarding binary gender. Values are adopted from Gabriel et al. (2008).

PARADIGM. The paradigm a given cosine similarity value belongs to is contained in PARADIGM. It takes the same value for all forms of a paradigm and was included as a random effect.

3.3 Results

The beta regression GAM revealed significant effects of COMPARISON and STEREOTYPICALITY. The effect of COMPARISON is displayed in Panel A of Figure 1. Bonferroni-corrected pairwise comparisons revealed that all possible comparisons are highly significant with one exception: generic masculines and specific feminines are just as similar as are specific masculines and gender star forms. Overall, the most similar forms are the generic and specific masculine. For STEREOTYPICALITY, it was found that higher values, i.e. paradigms that are stereotypically more female, come with lower cosine similarities values overall. This effect is shown in Panel B of Figure 1.

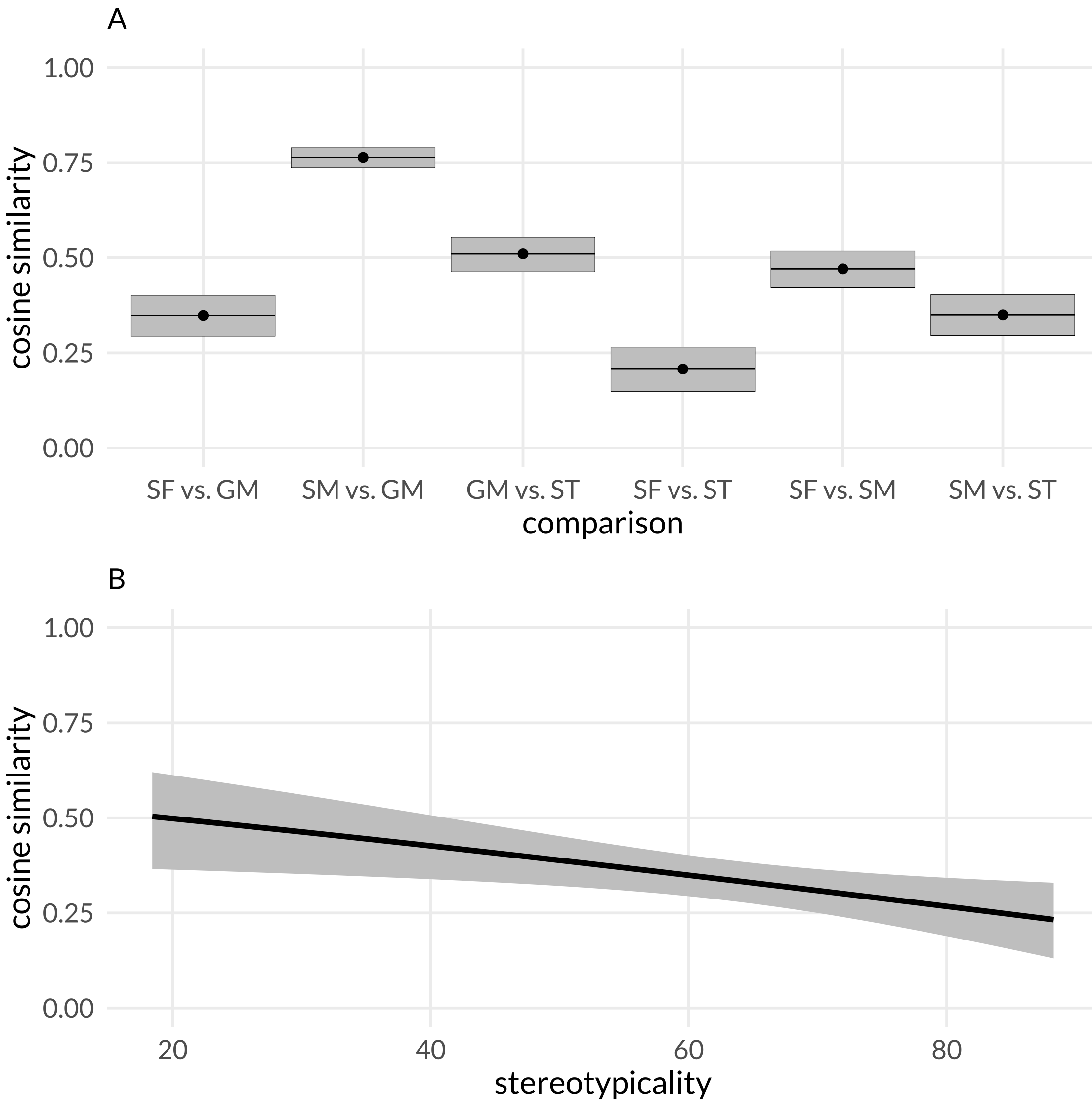

Figure 1: Mean cosine similarities as predicted by the beta regression GAM for COMPARISON (Panel A) and the partial effect of STEREOTYPICALITY (Panel B) with 95% confidence intervals.

3.4 Discussion

RQ1 asked how similar or dissimilar the semantics of generic masculines, gender star forms, and specific masculines and feminines are. Using cosine similarity as the measure of choice, the semantic vectors of all target paradigm forms were compared. For the two forms under investigation, it was found that generic masculines were most similar to specific masculines and that gender star forms were most similar to generic masculines. The high similarity of generic and specific masculines is in line with previous findings (e.g. Schmitz, Schneider & Esser 2023; Schmitz 2024) and provides further evidence for the male bias in generic masculines.

However, the present analysis of cosine similarities lacks detail. That is, it does not reflect potential effects of form frequencies, of form novelty, and of semantic interrelations in the mental lexicon. Further, semantic similarity is a rather crude measure, not allowing insight into why similarities differ in the way they do. An approach that does allow for detailed insights is discriminative learning.

4 Study 2: The gender star form from a discriminative learning perspective

4.1 Modelling comprehension: Linear discriminative learning

Linear discriminative learning, henceforth LDL, is the simple but interpretable machine learning implementation of the discriminative lexicon model, henceforth DLM, proposed by Baayen et al. (2019). The DLM is a computational theory of the mental lexicon, i.e. the representations and processes that together form the knowledge of words in an individual. In the DLM, comprehension and production are conceptualised as simple mappings between high-dimensional numerical representations of form and meaning, applying principles of error-driven learning. While the DLM is expandable by more complex non-linear features (cf. Chuang et al. 2024), it mostly works with simple linear networks. Keeping the mappings between form and meaning as simple as possible, the user is able to track how form and meaning relate to each other.

To obtain the meaning of a word form, the word form's vector representation c is mapped to its semantic vector via a comprehension transformation matrix F. As the involved vectors are generally of a high dimensionality, this mapping is not completely accurate. That is, the resulting semantic vector is a prediction of the actual semantics, and not virtually identical to the actual semantic vector. This is why the predicted semantic vector is called $\hat{s}$, not s. Similarly, in production, a word form's meaning s is mapped to its form vector via a production transformation matrix G, resulting in a predicted form $\hat{c}$.

Overall, there are three common ways to obtain the transformation matrices F and G (Heitmeier et al. 2024). First, multivariate regression modelling may be used to implement end-state learning. End-state learning, however, assumes infinite experience with the used language material, i.e. it may shed light on what is possible at the end of an infinite learning curve. Second, using the error-driven learning rule of Widrow and Hoff (Widrow & Hoff 1960), one may implement so-called incremental learning. Incremental learning uses information on learning order and word frequency, reflecting more closely the real learning process. Third, weighted versions of the form and semantic matrix may be used to implement so-called frequency informed learning (Heitmeier et al. 2024). Weighting makes use of word frequencies, while information on the learning order is not required. The weighted matrices are then made use of similarly to the non-weighted matrices in end-state learning. The results, however, are frequency-informed, which is reflected in overall better learning outcomes for high-frequency words as compared to overall worse learning outcomes for low-frequency words.

It is this latter way of obtaining the transformation matrix F that will be used in the present study. The following toy example illustrates frequency-informed learning more closely. In the form matrix C, each word form is allocated a number of rows based on its frequency. That is, $wordform1$ has a frequency of 2 and $wordform2$ has a frequency of 3:

$$C = \begin{array}{c} \\ wordform1 \\ wordform1 \\ wordform2 \\ wordform2 \\ wordform2 \end{array} \begin{array}{c} \begin{array}{ccc} cue1 & cue2 & cue3 \end{array} \\ \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix} \end{array} \tag{1}$$

To compute a frequency-informed transformation matrix F_f, we first must frequency-inform S and C to obtain S_f and C_f, as

$$F_f = {C_f}' S_f. \tag{2}$$

S_f and C_f are frequency-informed by a frequency-information matrix P. For the present example,

$$P = \begin{array}{c} wordform1 \\ wordform2 \end{array} \begin{pmatrix} \frac{2}{3} & 0 \\ 0 & \frac{3}{3} \end{pmatrix} \tag{3}$$

Using P, one can define $\tilde{C} = \sqrt{P}C$ and $\tilde{S} = \sqrt{P}S$. For instance,

$$\tilde{C} = \begin{array}{c} \\ wordform1 \\ wordform2 \end{array} \begin{array}{c} \begin{array}{ccc} cue1 & cue2 & cue3 \end{array} \\ \begin{pmatrix} \frac{2}{3} & \frac{2}{3} & 0 \\ 0 & \frac{3}{3} & \frac{3}{3} \end{pmatrix} \end{array}. \tag{4}$$

One can then calculate $\tilde{S} = \tilde{C}\tilde{F}$ and $\tilde{C} = \tilde{S}\tilde{G}$, respectively.[4]

Taking the original input matrices and the predicted matrices, a number of measures can be extracted for further analyses. For example, one may compare the observed and predicted vectors of a given word (as done in, e.g., Stein & Plag 2021). The closer these vectors are, for instance in terms of correlation coefficients, the more certain the comprehension or production process of the

[4] See Heitmeier et al. (2024) for further details on the mathematical workings of frequency-informed learning.

relevant word was. Another measure concerns the level of semantic co-activation of a given word: Conceptualised as, for example, Euclidean length, one may compare predicted word forms' meanings by their degree of caused co-activation in the mental lexicon (as done in, e.g., Schmitz, Schneider & Esser 2023). A third measure is the degree of uncertainty a given word form's meaning is subject to in the simulated mental lexicon (as done in, e.g., Schmitz et al. 2021). This measure is computed via ranked correlation coefficients of a given predicted vector and all observed vectors. The higher the overall correlation, the more uncertain the model is in the predicted vector, as there are many entries understood to show generally similar vectors, i.e. meanings or forms. Moving away from specific LDL measures, one may use the comprehended semantic vectors to perform an analysis of semantic similarity, making use of, for example, cosine similarity.

For the present paper, LDL was implemented based on frequency-informed learning. Words' forms were represented by trigrams, i.e. sequences of three grams. For example, the word *Haus* 'house' was represented by the trigrams `#ha`, `hau`, `aus`, and `us#`, where hash marks represent word boundaries. Trigrams were chosen over other forms of form representation, e.g. triphones and FBS features, as the present investigation is concerned with orthographic forms. Words' semantics were introduced as the semantic vectors computed via *fastText* (see Section 3.1). Frequencies for all 55,063 types were taken from the DWDS (Berlin-Brandenburgische Akademie der Wissenschaften 2024). For the grammatically masculine target words, obtaining their frequency included one extra step, as the DWDS frequency information does not differentiate between generic and specific masculines. Making use of the ratios between generic and specific masculines found by Schmitz, Schneider & Esser (2023) and the overall frequency of the masculines in the DWDS, the probable frequency of generic and specific masculines was calculated. As a toy example, assume that the masculine *Schüler* 'pupil' has a DWDS frequency of 12,000 and shows a generic usage ratio of 70%. For the frequency-informed learning, then, the frequency of generic *Schüler* is 8,400 and that of the specific *Schüler* is 3,600. For star forms, the frequency was set at that of their generic masculine counterpart, the reasoning of which will become clear shortly.

To gain detailed insight into the acquisition process of the novel star form and the comprehension connected to it, frequency-informed learning was not implemented at once but in 20 steps. During all steps, the frequencies of generic masculines, specific masculines, specific feminines, and non-target lexicon entries were kept constant. The frequencies of star forms, however, were increased randomly with each step, with null as initial frequency in the first step and their target frequency, i.e. that of their generic masculine counterpart, in the twentieth step. This step-wise frequency-based implementation allows insights into the trajectory of the comprehension of the gender-star forms as they become more frequent over time. That is, it simulates an increasing number of encounters with the gender-star form, just as it became more frequent in recent years. While there is no data available comparing the frequencies of generic masculines and gender star forms, putting their frequencies on a level in the final step nonetheless allows for an interesting, even if hypothetical, comparison. To obtain more reliable results, this step-wise frequency-informed LDL model was implemented 100 times, with each implementation using different random increases in gender star forms' frequencies.

4.2 Analysis

The similarity of the predicted semantics of the target word forms was analysed using cosine similarities. This analysis is overall similar to the one implemented in Section 3.2. That is, cosine similarities were introduced as the dependent variable in beta regression GAMs, again predicted by COMPARISON, STEREOTYPICALITY, and PARADIGM. Further, the following two variables were added:

RANDOMISATION. The number of the frequency randomisation a certain cosine similarity value belongs to is given in RANDOMISATION. The variable takes the values 1 to 100 and was included as a random effect.

STEP. The number of the step a given cosine similarity was taken from. The variable takes the values 1 to 20 and was included in interaction with COMPARISON, the motivation being that depending on the step, different comparisons might show different similarities.

The measures extracted for generic masculines and gender star forms from the LDL implementations were analysed with regular Gaussian GAMs using the same set of variables with one exception. That is, as the measures are not the result of a comparison, COMPARISON as a variable is not applicable. Instead, TYPE was introduced, encoding whether a given value is that of a generic masculine or gender star form. In the individual GAMs, TYPE is implemented in interaction with STEP, STEREO is given as additional predictor, and RANDOMISATION and PARADIGM are defined as random effects.

4.3 Results

First, let us take a look at the development of cosine similarities. Across all twenty models, COMPARISON showed highly significant effects. Taking a closer look at the intra-model differences between the different levels of COMPARISON using Bonferroni-corrected Tukey contrasts, it is found that only in a few instances comparisons are non-significant. These non-significant differences are given in Table 2. In other words, only in 8 cases, the overall similarity between two types is similar to the overall similarity between two other types. In step 2, specific feminines and generic masculines are as similar to each other as are specific feminines and gender star forms. This, however, is already different in step 3, where specific feminines and gender star forms are as similar to each other as generic masculines and gender star forms. The next non-significant difference is found in step 7 between specific feminines and gender star forms on the one hand, and specific masculines and gender star forms on the other. That is, at this step specific masculines and feminines are equally similar to gender star forms. The same is found in step 8. In steps 9 to 12, it is found that specific feminines and specific masculines are just as similar as are generic masculines and gender star forms.

Table 2: Non-significant differences found for COMPARISON across all models.

step	comparison	p-value
2	SF vs. GM and SF vs. ST	0.704
3	SF vs. ST and GM vs. ST	0.380
7	SF vs. ST and SM vs. ST	1.000
8	SF vs. ST and SM vs. ST	0.221
9	SF vs. SM and GM vs. ST	0.503
10	SF vs. SM and GM vs. ST	1.000
11	SF vs. SM and GM vs. ST	1.000
12	SF vs. SM and GM vs. ST	0.160

From step 13 on, all comparisons yield significantly different cosine similarities. Specific masculines and generic masculines are semantically most similar. A finding in line with the male bias found for generic masculines in previous computational (Schmitz 2023; Schmitz, Schneider & Esser 2023; Schmitz 2024) and many other studies (e.g. Stahlberg & Sczesny 2001; Körner et al. 2022; Schunack & Binanzer 2022; Zacharski & Ferstl 2023). The second most similar comparison is found for generic masculines and gender star forms, followed by specific feminines and specific masculines. Importantly, specific feminines are more similar to gender star forms than specific masculines are. This finding is in line with previous studies which found a feminine bias for gender star forms (e.g. Körner et al. 2022). The least semantically similar forms are specific feminines and generic masculines. The overall similarities across all steps are illustrated in Figure 2.

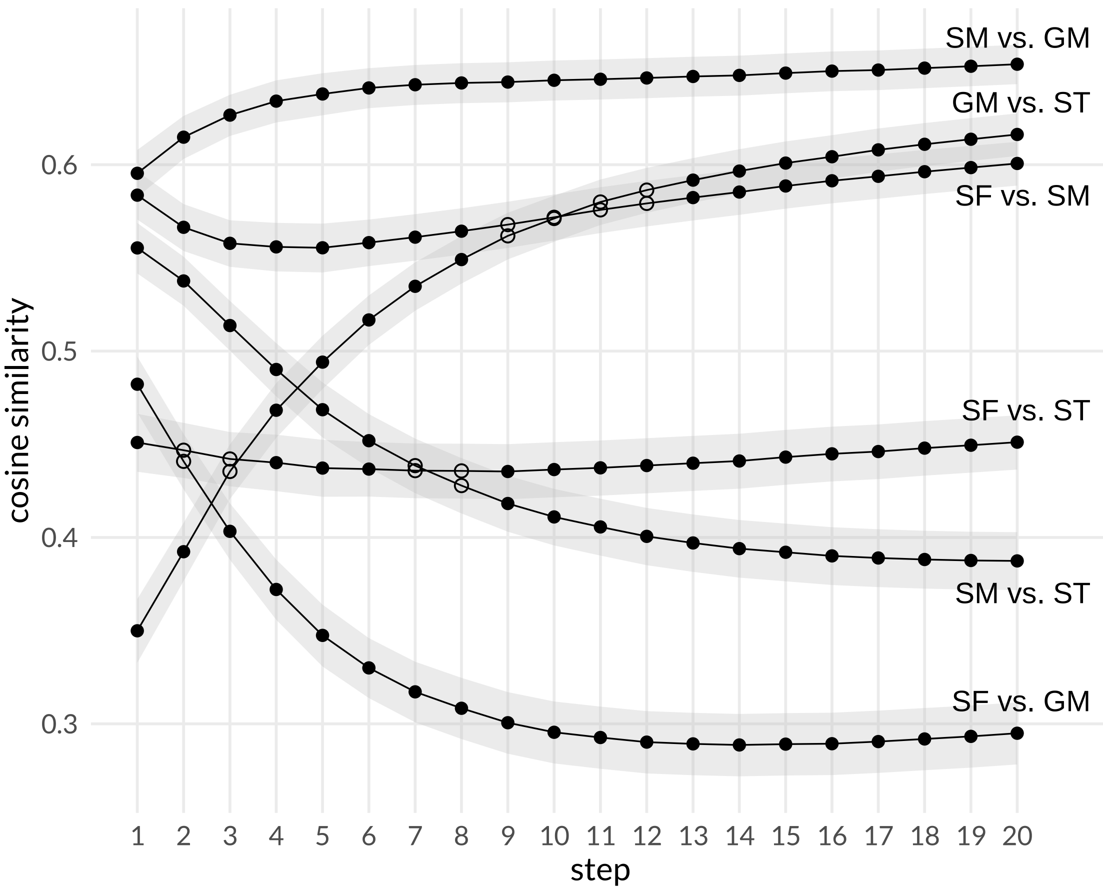

Figure 2: Mean predicted cosine similarities (points) for all comparisons across the 20 steps with 95% confidence intervals. Where points are empty, differences are non-significant, see Table 2.

Neither STEREOTYPICALITY nor RANDOMISATION reached significance in any model. PARADIGM, however, reached significance in all models, uncovering differences between the individual target word paradigms. The interpretation of these differences is beyond the scope of this paper.

While the analysis of semantic similarities and dissimilarities between different forms of a target word paradigm is interesting in itself, it lacks a direct link to the actual comprehension of language. That is, even if we assume that LDL reflects the lexical comprehension process and that this comprehension process is reflected within the semantic vectors predicted by the individual LDL implementations–and we do indeed assume that this is the case–relating further measures from the same LDL implementations to behavioural data offers the potential of further insight into the comprehension of target paradigm forms.

The three LDL measures used in the following analyses are the degree of semantic COACTIVATION, word-level CERTAINTY, i.e. the correlation of the original and the predicted semantic vector of a word, and lexicon-level UNCERTAINTY. These measures were introduced in Section 4.1, their interpretation is repeated here for reasons of convenience. COACTIVATION reflects the degree of semantic coactivation in the simulated mental lexicon when a given word is accessed. The higher its value, the higher the degree of semantic coactivation. That is, more semantic dimensions are more strongly coactivated. Word-level CERTAINTY reflects how well the comprehension process of a given word proceeded. The higher its value, the better a pertinent meaning was predicted by the LDL model. Lexicon-level UNCERTAINTY entails how uncertain the model was in predicting a given word form's semantics. The higher this value, the more uncertain the model was in a given predicted semantic vector. The difference between word-level CERTAINTY and lexicon-level UNCERTAINTY is that the latter takes into account the comprehension of all lexicon entries, while the earlier considers only a pertinent entry's comprehension. Figure 3 displays the three measures for generic masculines and gender star forms across all steps and randomisations. GAMs revealed that the differences between generic masculines and gender star forms are significant for all measures across all steps.[5]

[5] Model summaries can be found in the supplementary materials.

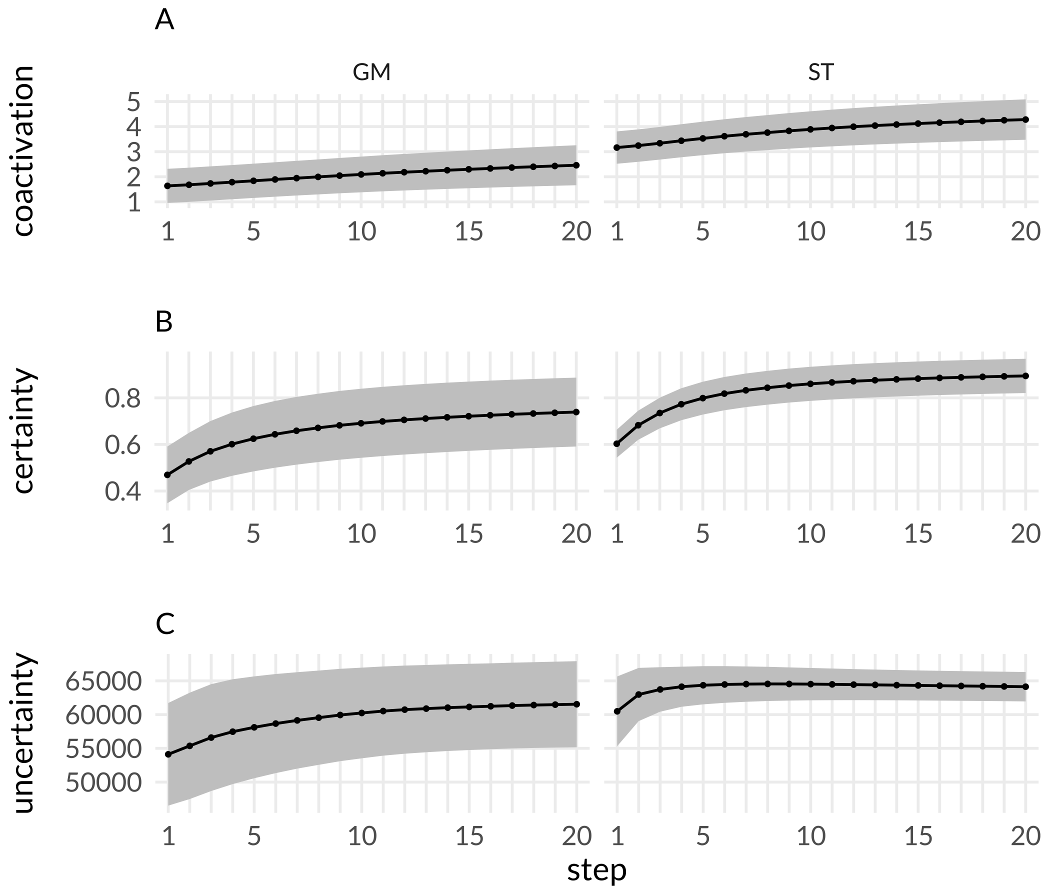

Figure 3: Mean observed values of semantic COACTIVATION (Panel A), word-level CERTAINTY (Panel B), and lexicon-level UNCERTAINTY (panel C) across all randomisations and LDL models with ±1 standard deviation.

For semantic COACTIVATION, generic masculines show a rising slope, starting roughly at 1.6 in the first step and ending with 2.5 in the final step. Star forms also show a rising slope, but already start higher than the maximum of the generic masculine, i.e. at circa 3.2, and ending at 4.3. Overall, gender star forms show the highest levels of semantic coactivation, indicating that more semantic dimensions are coactivated when a gender star form is accessed.

For word-level CERTAINTY, a steep rise during the first five steps can be seen for generic masculines, starting at circa 0.47 in the first step and ending at circa 0.62 in the fifth step. Going on, values keep increasing, reaching roughly 0.74 in the twentieth step. For gender star forms, a similar steep rise is found between steps 1 and 5. Starting at 0.60 already, the mean correlation coefficient at step 5 is approximately 0.80. Afterwards, the values keep growing, ending at circa 0.89 in step 20. Overall, the correlation between original and predicted vectors increases with increasing frequencies of gender star forms across steps. That is, the more familiar the simulated mental lexicon becomes with this novel form, the more certain it is not only in the gender star form itself, but also in the other non-novel but very much semantically related form of the generic masculine.

For lexicon-level UNCERTAINTY, one finds an overall increase of uncertainty from steps 1 to 20 for generic masculines, starting at circa 54,100 and ending at circa 61,500. Gender star forms show higher levels of uncertainty, starting at roughly 60,500 at step 1 and ending at roughly 64,100 at step 20. Notably, uncertainty rises during the first five steps, but plateaus afterwards and even decreases a little towards the end.

4.4 Discussion

Adding to the analysis of cosine similarities presented in Section 3, the semantic vectors predicted by 2000 implementations of LDL were analysed in a similar fashion. It was found that generic masculines were most similar to specific masculines across all steps. For gender star forms, similarity was higher with specific masculines than with specific feminines in early steps. From step 9 on, though, gender star forms were more similar to specific feminines than to specific masculines. That is, with increasing frequency , the semantics of gender star forms were closer to the semantics of specific feminines. This is a finding in line with previous research that found a female bias in gender star forms (e.g. Körner et al. 2022). Similarly, the overall low semantic similarity between generic masculines and specific feminines and the generally high similarity between generic masculines and specific masculines is in line with findings on the male bias of generic masculines (e.g. Schmitz, Schneider & Esser 2023). As for **RQ1**, it is found that even though the increase of gender star frequencies comes with recalibrations of similarities between different forms of role nouns, the highest semantic similarity is still found for generic and specific masculines.

To explain these difference further, semantic measures were extracted from LDL implementations, as **RQ2** asked how the semantics of generic masculines and gender star forms differ in detail. The three measures used were semantic COACTIVATION, word-level CERTAINTY, and lexicon-level UNCERTAINTY.

Gender star forms showed generally higher levels of semantic COACTIVATION than generic masculines, with increases in semantic COACTIVATION values for both forms across steps. In other words, gender star forms coactivate more semantic dimensions than generic masculines. One potential might lie in the gender-inclusive semantics of gender star forms: including more genders will potentially coactivate more semantic dimensions connected to these genders.

For word-level CERTAINTY, it was found that again gender star forms showed generally higher values as compared to generic masculines. For both forms, values of word-level CERTAINTY increase with increasing gender star frequencies. The LDL models were overall more certain in its predictions for gender star forms than for generic masculines. Again, one explanation for this finding might lie in their unique semantics, as they are the role nouns with strong relations to many semantic dimensions.

For lexicon-level UNCERTAINTY, gender star forms again showed generally higher values than generic masculines. Here, generic masculines showed an increase of lexicon-level UNCERTAINTY over increasing frequencies of gender star forms. Gender star forms only show an increase over the first few steps, and then a rather small decrease up to step 20. For generic masculines, the increasing levels of uncertainty across steps implies that the increasing frequency of the gender star forms does not help the interpretation of their semantics. One might argue that the gender star form, a form explicitly coined to represent all genders, renders the generic masculine, a form virtually identical to the specific masculine on the form level, even less gender-inclusive. The gender star form shows the highest levels of semantic uncertainty. One potential explanation might, again, lie in its rather strong connection to many semantic dimensions.

Lastly, one should address the seemingly contradictory findings for word-level CERTAINTY and lexicon-level UNCERTAINTY for gender star forms. How can word-level CERTAINTY imply that gender star forms come with the highest degree of certainty, while at the same time lexicon-level UNCERTAINTY implies that gender star forms come with the highest degree of uncertainty? Word-level CERTAINTY reflects how well the original semantics were predicted by the LDL model, while lexicon-level UNCERTAINTY reflects how many very similar vectors were predicted by the LDL model. That is, a predicted vector may be very close to its original counterpart, while at the same time there are many very similar vectors for other word forms. This, then, results in high values of word-level CERTAINTY and lexicon-level UNCERTAINTY. Hence, word-level CERTAINTY measures the certainty on the individual word level, while lexicon-level UNCERTAINTY measures the certainty of a given word form in relation to the entire simulated lexicon.

In the following section, behavioural data elicited in previous studies is reanalysed using the three measures extracted from the LDL implementations presented in the present section to reveal how these measures play out in language behaviour. Data were taken from Körner et al. (2022), Kurz & Mulder (2023), and Schunack & Binanzer (2022).

5 Study 3: Predicting language behaviour using measures from discriminative learning

5.1 Analysis

The analyses in this section make use of three types of models: beta regression in generalised additive mixed models, linear mixed-effects models, and cumulative link mixed-effects models. Whereas beta regression in GAMs was explained already in Section 3.2, the other two model types require an introduction.

Linear mixed-effects models were implemented for continuous dependent variables. While generalised additive mixed-models commonly retain all applicable variables, even those without significant effects, linear mixed-effects models are often reduced in a backward stepwise selection process (cf. Baayen 2008). This process starts out from a maximally specified model formula and checks in how far the removal of a given variable changes the overall model fit. If the removal of a given variable does not significantly increase model fit, the variable is dropped. If the removal of a variable leads to a significant decrease in model fit, the pertinent variable is retained. Linear mixed-effects models were fitted using the *lme4* package (Bates et al. 2015) and stepwise reduced using the *lmerTest* package (Kuznetsova, Brockhoff & Christensen 2017).

Cumulative link mixed-effects models were chosen for ordinal dependent variables and are overall similar to linear mixed-effects model. In contrast to linear mixed-effect models, though, no automated stepwise reduction function is available. Hence, the cumulative link mixed-effect models in this paper retained their initial formula specifications. Cumulative link mixed-effects models were fitted using the *ordinal* package (Christensen 2023).

After fitting a model, it was checked for issues of collinearity. Collinearity may lead to unreliable model estimates, rendering the model results meaningless (Amodio, Aria & D'Ambrosio 2014; Tomaschek, Hendrix & Baayen 2018). To check for such issues, variance inflation factors were used, where a value of > 10 is deemed problematic (James et al. 2013). In the case of such problematic variance inflation factors, pertinent variables were excluded and the model re-fitted.

In the following, for each set of data from a previous study, the original analysis from the data source will first be re-implemented. This is necessary as in all cases only a subset of the original data was used due to different sets of target items. Second, an updated version of the original model will be implemented, using additional and transformed variables. Third, the updated model will be re-implemented with the measures from the LDL implementations (see Section 4.3) replacing the categorical variable TYPE, which encodes whether a given form was a generic masculine or gender star form, as these measures represent such forms' semantic features. The motivation of this replacement is to uncover whether the LDL measures can provide further insight into the differences found between the categories of generic masculines and gender star forms presented by previous studies. Finally, the model fit of the original, the improved, and the model containing LDL measures are compared by their AIC values, revealing whether the inclusion of LDL measures not only provides insight into the underlying nature of generic masculines and gender star forms, but also indicating whether their inclusion actually benefits explaining the data.

5.2 Reaction times of continuation judgements

The data analysed from Körner et al. (2022) are the reaction time data taken from their second experiment (see Section 2.2 for a summary), as this data were elicited for generic masculines and gender star forms. Because the present LDL implementation did not cover all target words used in their experiment, a subset of the data (n = 8442) was analysed.

First, the original linear mixed-effects model from their supplementary material was recreated with the subset data. That is, reaction times were predicted by the TYPE of prime in the first sentence, i.e. a generic masculine or gender star form, in interaction with the gender of the referent in the CONTINUATION sentence, i.e. a male or female referent. Additionally, random intercepts for PARTICIPANTS with random slopes for CONTINUATION sentence referents' gender were included just as well as random intercepts for target word PARADIGMS with random slopes for the TYPE of prime. The results of this model are in line with those reported by Körner et al. (2022): only the interaction of TYPE and CONTINUATION reached significance. Bonferroni-corrected pairwise comparisons revealed that with generic masculines male continuations are judged faster and that with gender star forms female continuations are judged faster. Further, male continuations are judged significantly faster with generic masculines than with gender star forms in the first sentence.

Second, an updated version of the original model was created. Using the same subset of data, several variables were added to the model. Participants' AGE was included to control for effects of age on reaction times (cf. Fozard et al. 1994). The familiarity with generic masculines (GMFAM) and gender star (STFAM) forms, respectively, was added to account for potential influences of familiarity on the reaction times. The personal ATTITUDE towards gender-inclusive language was incorporated to take into account potential biases. Following other previous studies, the STEREOTYPICALITY of the individual paradigms was included using data from Gabriel et al. (2008). The last newly included variable was TRIALNUMBER, as reaction times may be subject to effects of task accommodation and fatigue. Further, reaction times were log-transformed to achieve a distribution closer to a normal distribution. After the exclusion of variables that do not contribute to a significantly better model fit, AGE and TRIALNUMBER were retained alongside the variables of the original model. Note, however, that the random slope for target word PARADIGMS by the TYPE of prime was dropped. The effects of the interaction of TYPE and CONTINUATION remained unchanged in nature, i.e. reaction times are faster for male continuations after generic masculines, for female continuations after gender star forms, and male continuations after generic masculines than after gender star forms. In addition, AGE and TRIALNUMBER showed significant effects. A higher age comes with slower reaction times; a later trial comes with faster reaction times.

Third, a version of the initial updated model was created with one important change: TYPE, i.e. whether the first sentence contained a target word in its generic masculine or gender star form, was replaced by the respective means of the three LDL measures COACTIVATION, word-level CERTAINTY, and lexicon-level UNCERTAINTY. As information on the familiarity with generic masculines and gender star forms was available, this information was matched by different samples of the three measures. That is, participants with the lowest level of STFAM were matched with measure values from the first step, whereas participants with the highest level STFAM were matched with values from the final step. While GMFAM could not be matched in a similar way, as the frequencies of generic masculines were not altered between the implementation steps, values were taken from the same step that values for gender star forms were taken from. Additionally, one might assume that even if participants were not aware of the concept of generic masculines, they nonetheless encounter them on a regular basis. After the exclusion of variables that do not contribute to a significantly better model fit, word-level CERTAINTY as main effect and COACTIVATION in interaction with CONTINUATION remained in the model alongside AGE, TRIALNUMBER, and the variables of the original model minus the random slope for target word PARADIGMs by the three LDL measures. Similar to the original interaction of TYPE and CONTINUATION, the interaction of COACTIVATION and CONTINUATION reached significance. AGE and TRIALNUMBER showed the same effects as in the previous model. An overview of the model is given in Table 3.

Table 3: Type III analysis of variance table with Satterthwaite's method for the final model including LDL measures.

	Sum Sq	Mean Sq	NumDF	DenDF	F-value	p-value	
COACTIVATION	0.283	0.283	1	589.6	3.405	0.065	
UNCERTAINTY	0.075	0.075	1	6784.5	0.901	0.343	
CONTINUATION	0.004	0.004	1	328.2	0.047	0.828	
COACTIVATION:CONTINUATION	3.504	3.504	1	1095.3	42.147	<0.001	***
AGE	0.847	0.847	1	365.8	10.181	0.002	**
TRIALNUMBER	90.664	90.664	1	7811.1	1090.400	<0.001	***

The effect of the interaction of COACTIVATION and CONTINUATION is illustrated in Figure 4. When the continuation sentence contained a female referent, reaction times did not change across the range of semantic COACTIVATION values. That is, irrespective of a pertinent form's semantic COACTIVATION, reaction times remained unchanged. When the continuation sentence, however, contained a male referent, reaction times changed significantly across the range of COACTIVATION values. With lower levels of semantic COACTIVATION, reaction times were fastest, while for higher levels of semantic COACTIVATION, reaction times were slowest.

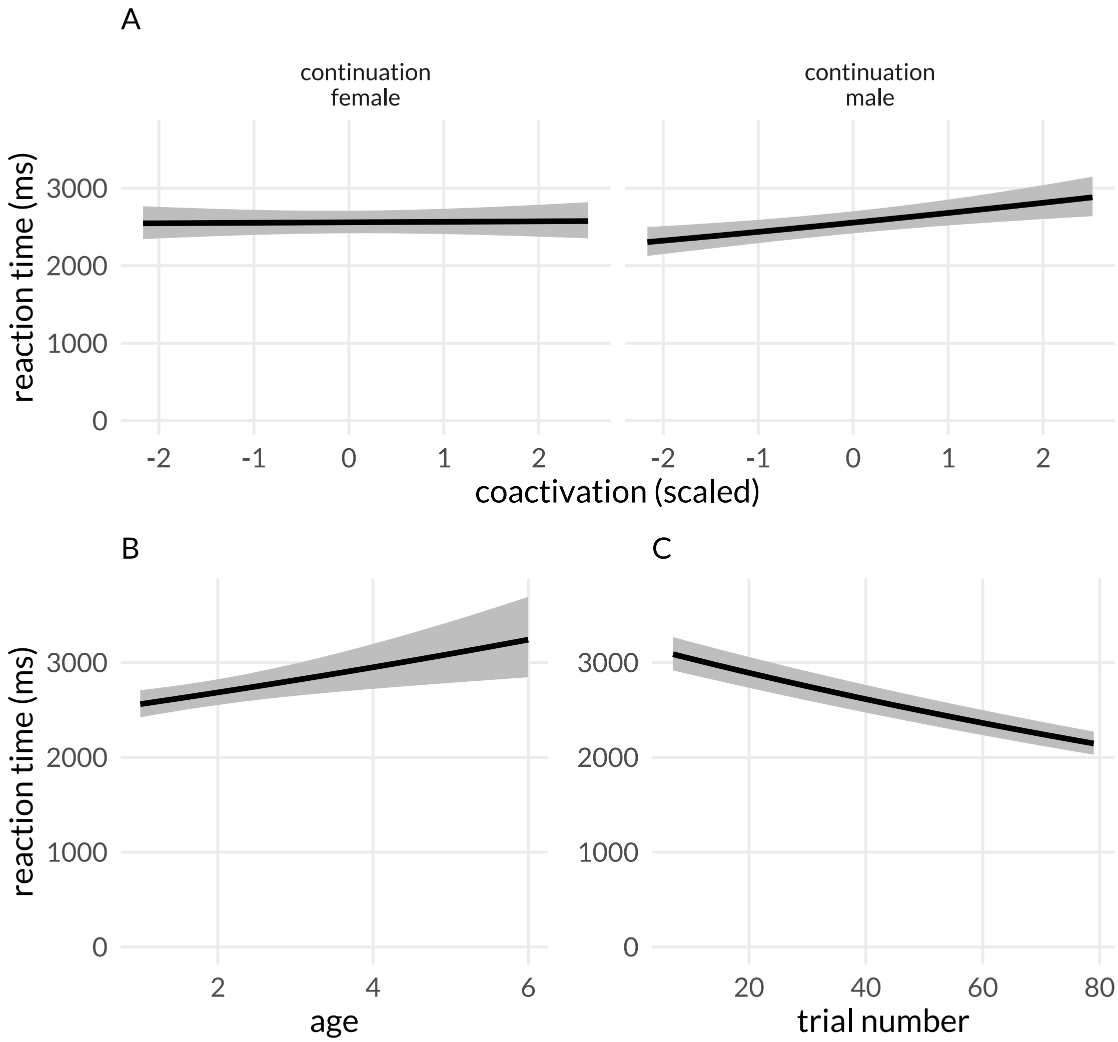

Figure 4: Partial effects of semantic COACTIVATION by CONTINUATION (Panel A), AGE (Panel B), and TRIALNUMBER (Panel C) on reaction time as predicted by the linear mixed-effects model with 95% confidence intervals.

Comparing the three linear mixed-effects models, the original model from the supplementary material by Körner et al. (2022) (AIC = 142644.813) was clearly outperformed by the updated model (AIC = 4439.006) and the model containing LDL measures instead of TYPE (AIC = 4435.014). The latter showed the overall best model fit.

For the reaction time data by Körner et al. (2022), it was found that when the continuation sentence contained a female referent, reaction times did not change across the range of semantic COACTIVATION values. That is, irrespective of a pertinent form's semantic COACTIVATION, reaction times remained unchanged. When the continuation sentence, however, contained a male referent, reaction times changed significantly across the range of coactivation values. With lower levels of semantic COACTIVATION, reaction times were fastest, while for higher levels of semantic COACTIVATION, reaction times were slowest. Considering that gender star forms come with generally higher levels of semantic COACTIVATION than generic masculines, one may conclude two things. First, male continuations are more difficult to judge after gender star forms than after generic masculines. This is in line with the results by Körner et al. (2022). Second, such judgements are apparently more difficult due to the higher degree of semantic COACTIVATION gender star forms are subject to. A higher degree of semantic COACTIVATION, then, may reflect that gender star forms do include semantics pertaining to more than one gender. This, in turn, might account for the null effect for female continuations: with forms of lower semantic COACTIVATION levels, i.e. generic masculines, a male bias is connected, while with forms of higher semantic COACTIVATION levels, i.e. gender star forms, more semantic dimensions are coactivated. Both the bias and the higher degree of semantic COACTIVATION exhibit about the same degree of influence on reaction times, rendering reaction times similar across the range of semantic COACTIVATION.

5.3 Proportion of male and female exemplars in exemplar recall

The proportion data taken from Kurz & Mulder (2023) (for a summary, see Section 2.2) were elicited for generic masculines and gender star forms of six target word paradigms: *Politiker vs. Politiker*innen* 'politicians', *Schauspieler vs. Schauspieler*innen* 'actors', *Autoren vs. Autor*innen* 'authors', *Athleten vs. Athlet*innen* 'athletes', *Sänger vs. Sänger*innen* 'singers', *TV-Moderatoren vs. TV-Moderator*innen* 'TV-hosts'. As *TV-Moderatoren vs. TV-Moderator*innen* was not part of the present LDL implementation, data on this target word paradigm were dropped, resulting in a subset of 495 data points.

First, the original analysis by Kurz & Mulder (2023) was re-implemented for the subset, that is a one-way ANOVA to predict the PROPORTION of female exemplars by the TYPE of target word presentation, i.e. generic masculine or gender star form. In line with the original findings, TYPE showed a highly significant effect on PROPORTION, with significantly fewer female exemplars for generic masculines.

Second, the original data was used to implement beta regression in generalised additive mixed models. As the proportional data contained zeros and ones, but beta regression cannot deal with true zeros and ones, zeroes were changed to 0.000001 and ones were changed to 0.999999. Proportions were predicted by TYPE, participant GENDER, target word paradigm STEREOTYPICALITY based on data by Gabriel et al. (2008), participant AGE, PARTICIPANT, and PARADIGM. The latter two were introduced as random effects. The model revealed significant effects for TYPE, as in the original

ANOVA, GENDER, and PARADIGM. Male participants provided more male exemplars than female participants; an effect also found by Kurz & Mulder (2023), even if only tentatively because of their modelling approach.

Third, a similar beta regression model was fitted with the important difference that TYPE was removed while the three LDL measures COACTIVATION, word-level CERTAINTY, and lexicon-level UNCERTAINTY were added. As there was no information available on the familiarity of participants with generic masculines and gender star forms, averages across all twenty steps were computed. These then should be somewhat representative of different levels of familiarity across participants. In the model with LDL measures, GENDER and PARADIGM again showed significant effects. All three LDL measures also influenced PROPORTION significantly, with the effects of COACTIVATION and word-level CERTAINTY being stronger than that of lexicon-level UNCERTAINTY ($p < 0.001$ vs. $p = 0.009$). AGE just reached significance with a p-value of 0.049. An overview of all effects is given in Table 4.

Table 4: Type III Wald tests of significance for the final model including LDL measures.

		df	Chi.sq	*p*-value	
GENDER		1	37.1	<0.001	***
	edf	Ref.df	Chi.sq	*p*-value	
COACTIVATION	1.000	1.000	32.335	<0.001	***
CERTAINTY	1.000	1.000	20.496	<0.001	***
UNCERTAINTY	1.134	1.244	13.059	0.009	**
STEREOTYPICALITY	1.000	1.000	0.003	0.954	
AGE	3.929	4.746	10.759	0.049	*
PARTICIPANT	8.473	96.000	9.669	0.149	
PARADIGM	2.889	3.000	73.061	<0.001	***

Figure 5 illustrates the significant effects. With higher levels of semantic COACTIVATION, more female exemplars are given as answers. For word-level CERTAINTY and lexicon-level UNCERTAINTY, lower values come with higher proportions of male exemplars. Female participants generally provide more female exemplars than male participants. The effect of AGE is inconclusive and should be taken with a grain of salt, as it barely reached significance.

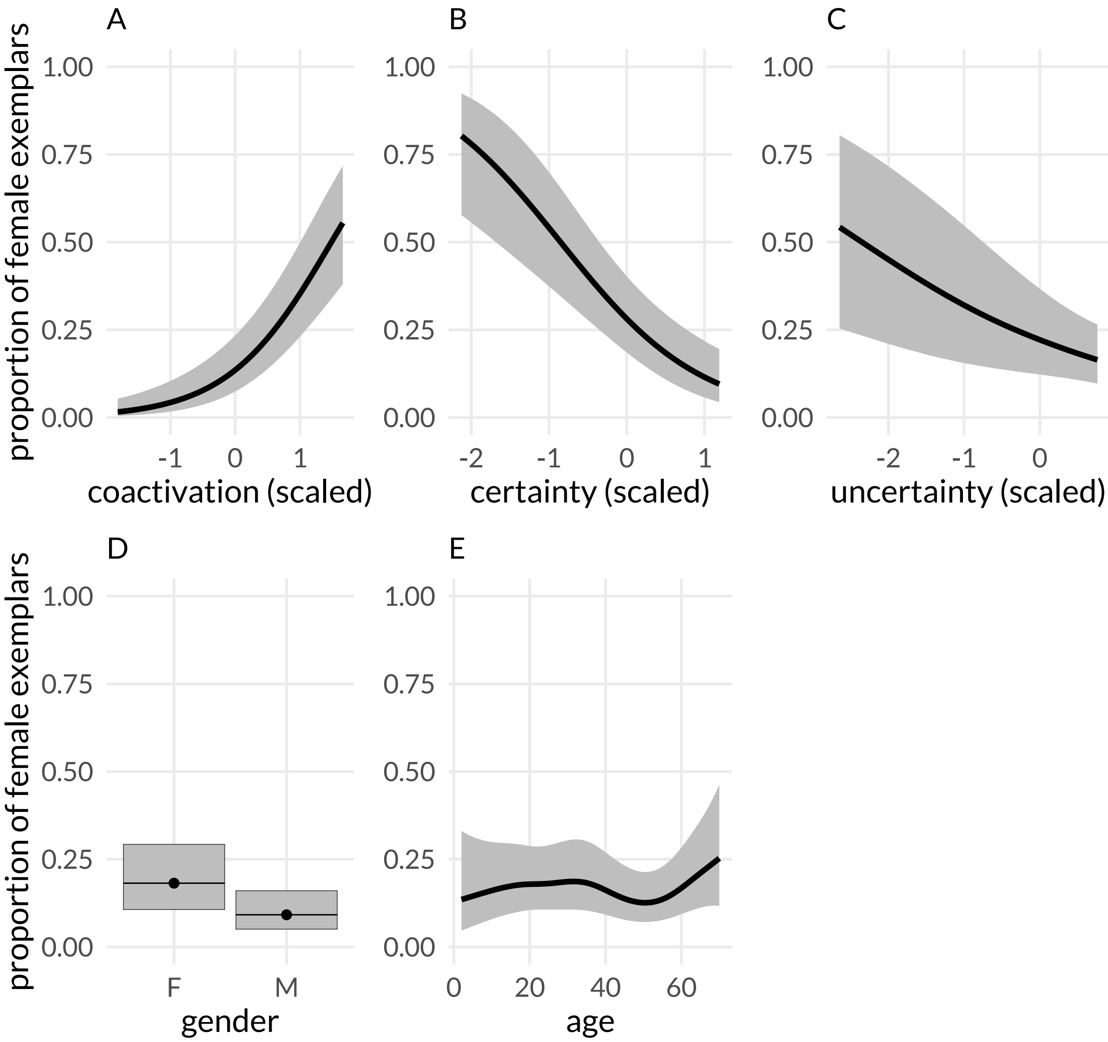

Figure 5: Partial effects of semantic COACTIVATION (Panel A), word-level CERTAINTY (Panel B), lexicon-level UNCERTAINTY (Panel C), GENDER (Panel D), and AGE (Panel E) on the proportion of female exemplars as predicted by the beta regression model with 95% confidence intervals.

Comparing the three analyses, the original analysis, i.e. the ANOVA, showed the overall worst model fit (AIC = 187.0182). Between the two beta regression models, the model containing the LDL measures (AIC = −9027.4492) outperformed the one using TYPE (AIC = −9002.4619).

For the proportion data by Kurz & Mulder (2023), all three LDL measures showed significant effects. With higher levels of semantic COACTIVATION, more female exemplars were given as answers. Taking into consideration that gender star forms come with generally higher levels of COACTIVATION, this is in line with the findings by Kurz & Mulder (2023) and implies that gender star forms more readily activate female exemplars. For word-level CERTAINTY and lexicon-level UNCERTAINTY, lower values come with higher proportions of female exemplars. That is, if a participant was asked to recall exemplars by being confronted with generic masculines, a lower word-level CERTAINTY came with higher proportions of female exemplars. Similarly, the same participant recalled more female exemplars when the generic masculine came with lower degrees of lexicon-level UNCERTAINTY. As generic masculines are semantically most similar to specific masculines, a high word-level CERTAINTY should come with higher proportions of male exemplars, just like a higher degree of lexicon-level UNCERTAINTY should do as well, as such a generic masculine would come with a semantically very similar specific masculine. If a participant was asked to recall exemplars by being confronted with gender star forms, a lower word-level CERTAINTY came with higher proportions of female exemplars. That is, if the novel form lead to word-level uncertainty, a female exemplar recall was more probable. One straightforward explanation for this finding lies in the form similarity of gender star and feminine forms. Further, the same participant recalled more female exemplars when the gender star form came with lower degrees of lexicon-level UNCERTAINTY. One potential explanation might lie in the apparent feminine bias in gender star form semantics, which would then come with the same consequences as the male bias in generic masculines.

5.4 Ratio of men vs. women in social and occupational groups

The ratio data taken from Schunack & Binanzer (2022) are from their second experiment (see Section 2.2 for a brief summary). As not all target word paradigms used in their experiment were part of the present LDL implementation, a subset of their data ($n = 2725$) was used. Additionally, data on capital I forms were dropped, as they, too, were not part of the LDL implementation.

First, the original analysis by Schunack & Binanzer (2022), a cumulative link mixed-effects model, was re-implemented using the subset data. The original ratio data was converted to an ordinal variable with values from 1 to 11. This rating scale was predicted by an interaction of paradigm member TYPE, i.e. generic masculine or gender star form, and STEREOTYPICALITYCATEGORY, i.e. female, male, or neutral. Additionally, random intercepts for PARTICIPANT and PARADIGM were specified. The main effects of TYPE and STEREOTYPICALITYCATEGORY reached significance, while their interaction did not. Gender star forms showed higher ratios of female group representatives than generic masculines, just like stereotypically feminine paradigms showed higher female ratios than stereotypically neutral paradigms, which in turn showed higher female ratios than stereotypically male paradigms.

Second, the original model was fitted but with additional variables. Participant AGE and binary GENDER were added as main effects, as well as participant FAMILIARITY with gender-inclusive language. Participant EDUCATION was specified as additional random intercept. The model results are in line with those of the re-implemented original model. The only newly emerged effect is that of FAMILIARITY: somewhat surprisingly, participants more familiar with gender-inclusive language showed lower ratios of women in their ratings.

Third, the model with additional variables was re-fitted but with the three LDL measures COACTIVATION, word-level CERTAINTY, and lexicon-level UNCERTAINTY in interaction with STEREOTYPICALITYCATEGORY instead of TYPE. As the FAMILIARITY measure is rather coarse, it is binary with either `yes` or `no` as value, LDL measures were implemented as means across steps. For participants unfamiliar with gender star forms, the mean of steps 1 to 5 was used, while for participants familiar with gender star forms, the mean of steps 15 to 20 was used. Interestingly, now that TYPE was replaced with the three LDL measures that describe the type of paradigm member more closely, the interaction of two of these measures, COACTIVATION and word-level CERTAINTY, with STEREOTYPICALITYCATEGORY reached significance. Table 5 provides a model summary.

Table 5: Summary of the final model including LDL measures.

	Estimate	Std. Error	z-value	p-value	
COACTIVATION	-0.422	0.207	-2.040	0.041	*
STEREOCAT`masc`	-7.325	0.297	-24.703	<0.001	***
STEREOCAT`neut`	-4.329	0.243	-17.808	<0.001	***
CERTAINTY	0.961	0.222	4.329	<0.001	***
UNCERTAINTY	0.018	0.107	0.170	0.865	
GENDER`male`	-0.106	0.087	-1.212	0.226	
AGE	-0.006	0.005	-1.144	0.252	
FAMILIARITY`yes`	-0.006	0.120	-0.047	0.962	
COACTIVATION:STEREOCAT`masc`	0.632	0.224	2.818	0.005	**
COACTIVATION:STEREOCAT`neut`	0.483	0.215	2.251	0.024	*
CERTAINTY:STEREOCAT`masc`	-1.263	0.233	-5.411	<0.001	***
CERTAINTY:STEREOCAT`neut`	-0.894	0.222	-4.021	<0.001	***
UNCERTAINTY:STEREOCAT`masc`	-0.068	0.145	-0.470	0.639	
UNCERTAINTY:STEREOCAT`neut`	-0.039	0.126	-0.309	0.757	

The effects of both significant interactions are given in Figure 6.[6] For stereotypically feminine target paradigms, a decrease in the ratio of women is found with increasing values of semantic COACTIVATION, while for stereotypically male target paradigms the opposite is the case. For word-level CERTAINTY, stereotypically feminine target paradigms show an increase in the ratio of women

[6]Note that this figure was created using a linear mixed-effects model following the same formula as the main cumulative link mixed-effects model, as to the author's knowledge there is no plotting option for the latter available. All effects in the linear mixed-effects model were similar to those in the cumulative link mixed-effects model.

and stereotypically male target paradigms show an increase in the ratio of men with increasing values of word-level CERTAINTY. For stereotypically neutral paradigms, no differences across the range of word-level CERTAINTY is found.

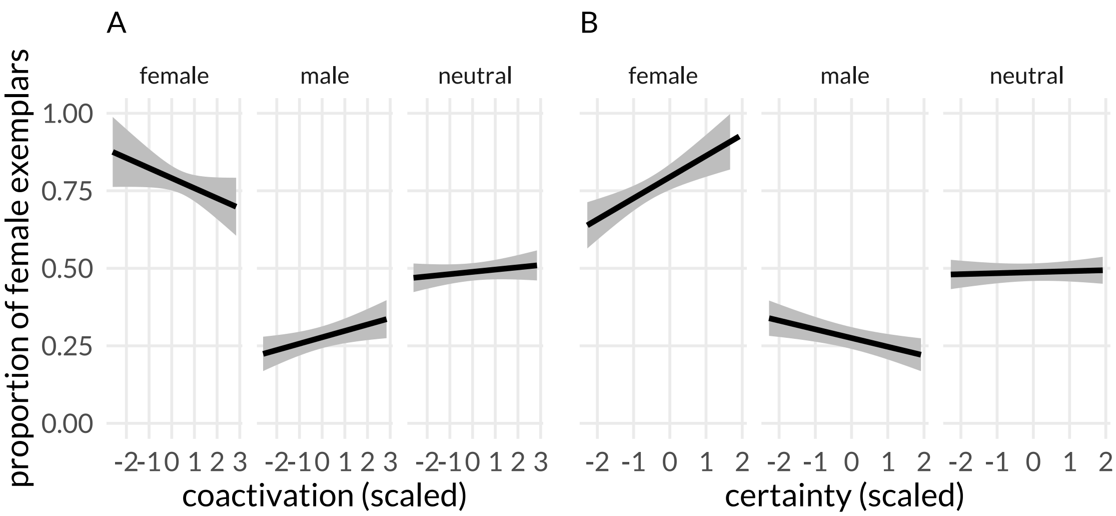

Figure 6: Partial effects of the interactions of semantic COACTIVATION (Panel A) and word-level CERTAINTY (Panel B) with STEREOTYPICALITYCATEGORY with 95% confidence intervals.

As for model fit, the original cumulate link mixed-effects model showed the worst fit (AIC = 9776.037), followed by its updated version (AIC = 9773.540). The model containing the three LDL measures showed the overall best model fit (AIC = 9741.412).

For the ratio data by Schunack & Binanzer (2022), semantic COACTIVATION and word-level CERTAINTY reached significance. For stereotypically feminine target paradigms, a decrease in the ratio of women was found with increasing values of semantic COACTIVATION, while for stereotypically male target paradigms, an increase in the ratio of women was found with increasing values of semantic COACTIVATION. That is, gender star forms, which come with generally higher levels of semantic coactivation, lead to a decrease in the ratio of women in stereotypically female and to an increase in the ratio of women in stereotypically male groups. Hence, the gender star form works against the stereotypicality of such groups. For word-level CERTAINTY, however, one finds an effect of a somewhat different nature. Stereotypically feminine target paradigms show an increase in the ratio of women with increasing values of word-level CERTAINTY. The closer the predicted vector is to its observed counterpart, i.e. the more certain the model is in its prediction, the higher the ratio of women in paradigms which already show a generally high ratio of women due to their stereotypicality. Similarly, the ratio of women decreases with generic masculines of a high word-level CERTAINTY value. For stereotypically neutral paradigms, no differences across the range of word-level CERTAINTY were found. It appears that semantic COACTIVATION and word-level CERTAINTY show effects of opposing nature.

5.5 Discussion

RQ3 asked which semantic measures account for the differences found between generic masculines and gender star forms in previous studies. By combining the three measures semantic COACTIVATION, word-level CERTAINTY, and lexicon-level UNCERTAINTY with data elicited in previous studies on generic masculines and gender star forms, it was found that, depending on the experimental task, different measures account for the differences found between generic masculines and gender star forms. That is, measures extracted from implementations of LDL allow a more detailed explanation of why the two forms lead to different reaction times in continuation judgements, to different proportions of male and female exemplars in exemplar recall, and to different ratios of men vs. women in social and occupational groups.

6 General discussion

The present study set out to provide a first account of gender star form semantics and to explain the differences between generic masculines and gender star forms found in previous studies.

First, **RQ1** asked how similar or dissimilar the semantics of generic masculines, gender star forms, and specific masculines and feminines are. Using cosine similarity to assess the similarity between vector representations of target words computed with *fastText*, it was found that generic masculines are most similar to specific masculines. A finding in line with previous computational studies (Schmitz, Schneider & Esser 2023; Schmitz 2024). For gender star forms, it was found that they are most similar to generic masculines, followed by specific masculines. However, as this account of semantic similarity did not reflect potential influences of form frequency, form novelty, and interrelations in the mental lexicon, a second similar but advanced analysis was conducted. Using LDL, frequency-informed networks were implemented in which the frequency of gender star forms increased across 20 steps. To account for the random nature of potential real frequency increases, 100 different series of random frequency increases were used. Analysing the predicted semantic vectors of role nouns across the 2,000 resulting LDL models, it was again found that generic masculines and specific masculines are semantically most similar. For gender star forms, it was found that again they are more similar to specific masculines than to specific feminines, but only until their frequency is roughly half of that of generic masculines. Once their frequency is higher than this threshold, they become semantically more similar to specific feminines. This finding, then, is in line with previous research that found a female bias in gender star forms (e.g. Körner et al. 2022).

Moving on from this rather broad description of semantic similarities and dissimilarities, **RQ2** asked how the semantics of generic masculines and gender star forms differed in detail. Using the aforementioned implementations of LDL, three semantic measures were computed: semantic COACTIVATION, word-level CERTAINTY, and lexicon-level UNCERTAINTY. Gender star forms showed generally higher levels of all three measures than generic masculines. That is, gender star forms coactivate more semantic dimensions when they are comprehended, they come with a higher degree of word-level certainty, and they come with a higher degree of lexicon-level uncertainty.

For the reanalysis of linguistic data elicited in previous studies and a question to **RQ3**, these differences in semantic measures allowed more detailed insights into the differences found between generic masculines and gender star forms. For reaction time data of continuation judgements, it was found that it is moderated by semantic COACTIVATION. That is, when confronted with a stereotypically male target, the high semantic COACTIVATION of gender star forms leads to slower reaction times, whereas for stereotypically female targets it does not. Here, however, the overall lower levels of semantic COACTIVATION of generic masculines do not lead to faster reaction times, as apparently their semantic bias works against the stereotypicality of the target. In the recall of female exemplars, the higher levels of semantic COACTIVATION in gender star forms come with higher proportions of female exemplars. Finally, in assessing the ratio of men vs. women in social and occupational groups, the high level of semantic COACTIVATION in gender star forms leads to a lower proportion of women in stereotypically feminine groups and to a lower proportion of men in stereotypically male groups. Overall, these findings point towards a more gender-inclusive semantic nature of gender star forms.

Word-level CERTAINTY and lexicon-level UNCERTAINTY also accounted for previous findings. That is, in exemplar recall, lower values of both measures came with higher proportions of female exemplars. If a participant was asked to recall exemplars via a generic masculine, a lower word-level CERTAINTY and a lower degree of lexicon-level UNCERTAINTY came with higher proportions of female exemplars. Due to their semantic similarity with specific masculines, high degrees of the two measures should come with higher proportions of male exemplars. If a participant was asked to recall exemplars via a gender star form, a lower word-level CERTAINTY and a lower degree of lexicon-level UNCERTAINTY came with higher proportions of female exemplars. In other words, if the novel gender star form lead to more word-level CERTAINTY, a female exemplar recall was more probable. A potential explanation for this finding lies in the rather similar forms of gender stars and feminines. A potential explanation for the effect of lexicon-level UNCERTAINTY lies in the feminine bias in gender star forms, which leads with the same consequences as the male bias in generic masculines. Finally, word-level CERTAINTY may also increase the ratio of the gender that is encoded in a target's stereotypicality. With stereotypically feminine target words, a higher ratio of women in a group was found with higher levels of word-level CERTAINTY, while with stereotypically male target words, a higher ratio of men in a group was found with higher levels of word-level CERTAINTY.

Finally, one issue remains unanswered: Do gender star forms effectively represent genders beyond the binary as intended? With the present data and analyses, this question cannot be answered. Nonetheless, this issue raises several questions. First, how would one check non-binary semantics using distributional semantics? Potentially, one would assume equal similarities between specific masculines, specific feminines, and the gender star form. This could, in theory, also indicate that the gender star form represents both binary genders equally well, with non-binary genders not being included. Another potential solution might lie in comparing the three forms to words with lexico-semantic gender information. However, there are no such words for non-binary genders. Second, even if the first issue was resolvable, non-binary individuals are less represented in text corpora and, for that matter, most written media. This is only somewhat surprising, as there are fewer non-binary than binary individuals in the public eye. Hence, one would need to find a suitable source to create a text corpus, as such a corpus is required to compute reliable vector representations of non-binary representing words. Third, even if the prior two issues were resolved, re-analyses as were done in the present paper are, with the current body of linguistic research, not applicable. That is, for example, exemplars of non-binary genders are nearly non-existent in studies, even though the gender star form invites the recall of non-binary individuals. Again, this is little surprising due to the ratio of binary vs. non-binary individuals in the public. Hence, it appears that this issue is not resolvable, at least not with our present resources.

7 Conclusion

Overall, both forms – generic masculines and gender star forms – show a semantic bias. Generic masculines are biased towards male readings, while gender star forms are biased towards feminine readings. Both forms show differences in their underlying semantic features: Generic masculines coactivate fewer semantic dimensions, are harder to predict, and exhibit less uncertainty for the comprehension process. Gender star forms, on the contrary, coactivate more semantic dimensions, are easier to predict, and lead to more uncertainty during comprehension.

These differences in their general semantics and in their semantic features account for differences found in previous studies. Higher levels of semantic coactivation lead to slower reaction times for gender star forms, but also to increased shares of female exemplars. The high similarity between specific and generic masculines and the rather low levels of comprehension and semantic certainty of the latter benefit the recall of male exemplars.

Future investigations should consider four main routes. First, a more qualitative approach to the present findings is required. For example, do all target word paradigms show the same similarities and differences of cosine similarity and LDL measures between their different forms as they were presented in this paper? Second, one should take into account further semantic measures, e.g. semantic neighbourhood density, but also phonological measures pertaining to the production process. Third, LDL implementations with form representations based on phonology or phonetics instead of orthography are called for, as the novel suffixes are also part of spoken language. Fourth, novel ways need to be found to include non-binary gender representations in semantic analyses of gender star and other gender-inclusive forms.

In sum, the current study presented three novel accounts. First, an account of the semantics of gender star forms and the relations of these semantics to those of the specific masculine, specific feminine, and generic masculine. Second, the first computational account of these semantics using linear discriminative learning, the computational engine of the discriminative lexicon. Third, measures extracted from the computational analysis were used to re-analyse data from previous studies, providing insights into how these measures account for differences in reaction times for continuation judgements, proportions of male and female exemplars in exemplar recall, and ratios of men vs. women in social and occupational groups.

8 Supplementary Material

Find all data and scripts at https://osf.io/gaxsq/?view_only=9ab7007e57fc49629884d5d8d3e2e920.

References

‚Tagesspiegel' schafft Gendersternchen ab. 2023. *Frankfurter Allgemeine*. https://www.faz.net/aktuell/feuilleton/medien/tagesspiegel-schafft-gendersternchen-nach-beschwerden-ab-19345892.html (17 May, 2024).

Amodio, Sonia, Massimo Aria & Antonio D'Ambrosio. 2014. On concurvity in nonlinear and nonparametric regression models. *Statistica* 74. 85–98. https://doi.org/10.6092/issn.1973-2201/4599.

Baayen, R. Harald. 2008. *Analyzing linguistic data: A practical introduction to statistics using R*. 1–353. https://doi.org/10.1017/CBO9780511801686.

Baayen, R. Harald, Yu-Ying Chuang, Elnaz Shafaei-Bajestan & James P. Blevins. 2019. The discriminative lexicon: A unified computational model for the lexicon and lexical processing in comprehension and production grounded not in (de)composition but in linear discriminative learning. *Complexity* 2019. 4895891. https://doi.org/10.1155/2019/4895891.

Baayen, R. Harald, Petar Milin, Dusica Filipović Đurđević, Peter Hendrix & Marco Marelli. 2011. An amorphous model for morphological processing in visual comprehension based on naive discriminative learning. *Psychological Review* 118. 438–481. https://doi.org/10.1037/a0023851.

Bates, Douglas, Martin Mächler, Benjamin M. Bolker & Steven C. Walker. 2015. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67. https://doi.org/10.18637/jss.v067.i01.

Berlin-Brandenburgische Akademie der Wissenschaften. 2024. *DWDS – Digitales Wörterbuch der deutschen Sprache*. (17 May, 2024).

Bojanowski, Piotr, Edouard Grave, Armand Joulin & Tomas Mikolov. 2016. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics* 5. 135–146. https://doi.org/10.48550/arxiv.1607.04606.

Boleda, Gemma. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6. 213–234. https://doi.org/10.1146/annurev-linguistics-011619-030303.

Braun, Friederike, Anja Gottburgsen, Sabine Sczesny & Dagmar Stahlberg. 1998. Können Geophysiker Frauen sein? Generische Personenbezeichnungen im Deutschen. *Zeitschrift für Germanistische Linguistik* 26. 265–283. https://doi.org/10.1515/zfgl.1998.26.3.265.

Christensen, Rune H. B. 2023. *ordinal—Regression Models for Ordinal Data*. R package version 2023.12-4. https://CRAN.R-project.org/package=ordinal.

Chuang, Yu-Ying, Melanie J. Bell, Y.-H. Tseng & R. Harald Baayen. 2024. Word-specific tonal realizations in mandarin. *Preprint*.

Diewald, Gabriele. 2018. Zur Diskussion: Geschlechtergerechte Sprache als Thema der germanistischen Linguistik – exemplarisch exerziert am Streit um das sogenannte generische Maskulinum. *Zeitschrift für germanistische Linguistik* 46. 283–299.

Doleschal, Ursula. 2002. Das generische Maskulinum im Deutschen. Ein historischer Spaziergang durch die deutsche Grammatikschreibung von der Renaissance bis zur Postmoderne. *Linguistik Online* 11. https://doi.org/10.13092/lo.11.915.

Fozard, James L., Max Vercruyssen, Sara L. Reynolds, P. A. Hancock & Reginald E. Quilter. 1994. Age differences and changes in reaction time: the baltimore longitudinal study of aging. *Journal of Gerontology* 49. 179–189. https://doi.org/10.1093/geronj/49.4.P179.

Gabriel, Ute, Pascal Gygax, Oriane Sarrasin, Alan Garnham & Jane Oakhill. 2008. Au pairs are rarely male: Norms on the gender perception of role names across English, French, and German. *Behavior Research Methods* 40. 206–212. https://doi.org/10.3758/BRM.40.1.206.

Goldhahn, Dirk, Thomas Eckart & Uwe Quasthoff. 2012. Building large monolingual dictionaries at the Leipzig Corpora Collection: From 100 to 200 languages. In *Proceedings of the 8th International Language Resources and Evaluation (LREC'12)*.

Gygax, Pascal, Daniel Elmiger, Sandrine Zufferey, Alan Garnham, Sabine Sczesny, Lisa von Stockhausen, Friederike Braun & Jane Oakhill. 2019. A language index of grammatical gender dimensions to study the impact of grammatical gender on the way we perceive women and men. *Frontiers in Psychology* 10. https://doi.org/10.3389/fpsyg.2019.01604.

Gygax, Pascal, Ute Gabriel, Oriane Sarrasin, Jane Oakhill & Alan Garnham. 2008. Generically intended, but specifically interpreted: When beauticians, musicians, and mechanics are all men. *Language and Cognitive Processes* 23. 464–485. https://doi.org/10.1080/01690960701702035.

Hanitzsch, Thomas. 2021. Genderstern und Binnen-I: Es ist Zeit, die Realität zu akzeptieren. *Publizistik 2021 66:2* 66. 181–185. https://doi.org/10.1007/S11616-021-00645-4.

Harris, Zellig S. 1954. Distributional structure. *WORD* 10. 146–162. https://doi.org/10.1080/00437956.1954.11659520.

Heitmeier, Maria, Yu-Ying Chuang, Seth D. Axen & R. Harald Baayen. 2024. Frequency effects in linear discriminative learning. *Frontiers in Human Neuroscience* 17. https://doi.org/10.3389/fnhum.2023.1242720.

Irmen, Lisa & Julia Kurovskaja. 2010. On the semantic content of grammatical gender and its impact on the representation of human referents. *Experimental Psychology* 57. 367–375. https://doi.org/10.1027/1618-3169/a000044.

James, Gareth, Daniela Witten, Trevor Hastie & Robert Tibshirani (eds.). 2013. *An introduction to statistical learning: With applications in R*. New York: Springer.

Jarnatowskaja, V. E. 1968. Die Kategorie des Genus der Substantive im System der deutschen Gegenwartssprache. *DaF* 5. 213–219.

Keith, Nina, Kristine Hartwig & Tobias Richter. 2022. Ladies first or ladies last: Do masculine generics evoke a reduced and later retrieval of female exemplars? *Collabra: Psychology* 8. https://doi.org/10.1525/COLLABRA.32964/120263.

Körner, Anita, Bleen Abraham, Ralf Rummer & Fritz Strack. 2022. Gender representations elicited by the gender star form. *Journal of Language and Social Psychology* 41. 553–571. https://doi.org/10.1177/0261927X221080181.

Körner, Anita, Sarah Glim & Ralf Rummer. 2024. Examining the glottal stop as a mark of gender-inclusive language in german. *Applied Psycholinguistics*. 1–24. https://doi.org/10.1017/S0142716424000018.

Kurz, Pia & Hannah De Mulder. 2023. A star is born? The German gender star and its effects on mental representation. *Psychology of Language and Communication* 27. 384–404. https://doi.org/10.58734/PLC-2023-0018.

Kuznetsova, Alexandra, Per B. Brockhoff & Rune H. B. Christensen. 2017. lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82. https://doi.org/10.18637/jss.v082.i13.

Mikolov, Tomas, Kai Chen, Greg Corrado & Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *1st International Conference on Learning Representations, ICLR 2013 - Workshop Track Proceedings*. https://doi.org/10.48550/arxiv.1301.3781.

Misersky, Julia, Pascal M. Gygax, Paolo Canal, Ute Gabriel, Alan Garnham, Friederike Braun, Tania Chiarini, Kjellrun Englund, Adriana Hanulikova, Anton Öttl, Jana Valdrova, Lisa Von Stockhausen & Sabine Sczesny. 2014. Norms on the gender perception of role nouns in Czech, English, French, German, Italian, Norwegian, and Slovak. *Behavior Research Methods* 46. 841–871. https://doi.org/10.3758/S13428-013-0409-Z.

Misersky, Julia, Asifa Majid & Tineke M. Snijders. 2019. Grammatical gender in German influences how role-nouns are interpreted: Evidence from ERPs. *Discourse Processes* 56. 643–654. https://doi.org/10.1080/0163853X.2018.1541382.

Pennington, Jeffrey, Richard Socher & Christopher D. Manning. 2014. *GloVe: Global vectors for word representation*. nlp.stanford.edu/projects/glove/.

Pusch, Luise. 1984. *Das Deutsche als Männersprache*. Suhrkamp Verlag.

Python Software Foundation. 2023. *Python Language Reference, version 3.10.12*. http://www.python.org.

Řehůřek, Radim & Petr Sojka. 2010. Software framework for topic modelling with large corpora. English. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*, 45–50. Valletta, Malta: ELRA. http://is.muni.cz/publication/884893/en.

Rothermund, K. 1998. Automatische geschlechtsspezifische Assoziationen beim Lesen von Texten mit geschlechtseindeutigen und generisch maskulinen Text-Subjekten. *Sprache & Kognition* 17. 183–198.

Rothmund, Jutta & Brigitte Scheele. 2004. Personenbezeichnungsmodelle auf dem Prüfstand. *Zeitschrift für Psychologie / Journal of Psychology* 212. 40–54. https://doi.org/10.1026/0044-3409.212.1.40.

Schmid, Helmut. 1999. Improvements in part-of-speech tagging with an application to German. In S. Armstrong, K. Church, P. Isabelle, S. Manzi, E. Tzoukermann & D. Yarowsky (eds.), *Natural language processing using very large corpora*, 13–25. Springer. https://doi.org/10.1007/978-94-017-2390-9_2.

Schmitz, Dominic. 2023. In German, all professors are male. In Jasmin Pfeifer, Sabine Arndt-Lappe, Heidrun Dorgeloh, Gero Kunter & Christian Uffmann (eds.), *Ingo 6.0. The proceedings. New empirical Insights about laNguage, presented on a Great day Out in September*. Preprint. https://doi.org/10.31234/osf.io/yjuhc.

Schmitz, Dominic. 2024. Instances of bias: The gendered semantics of generic masculines in German revealed by instance vectors. *Zeitschrift für Sprachwissenschaft*.

Schmitz, Dominic, Ingo Plag, Dinah Baer-Henney & Simon David Stein. 2021. Durational differences of word-final /s/ emerge from the lexicon: Modelling morpho-phonetic effects in pseudowords with linear discriminative learning. *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.680889.

Schmitz, Dominic, Viktoria Schneider & Janina Esser. 2023. No genericity in sight: An exploration of the semantics of masculine generics in German. *Glossa Psycholinguistics* 2. https://doi.org/10.5070/G6011192.

Schunack, Silke & Anja Binanzer. 2022. Revisiting gender-fair language and stereotypes – a comparison of word pairs, capital I forms and the asterisk. *Zeitschrift für Sprachwissenschaft* 41. 309–337. https://doi.org/10.1515/ZFS-2022-2008.

Stahlberg, Dagmar, Friederike Braun, Lisa Irmen & Sabine Sczesny. 2007. Representation of the sexes in language. In A. W. Kruglanski & J. P. Forgas (eds.), *Social communication*, 163–187. New York: Psychology Press. https://doi.org/10.4324/9780203837702.

Stahlberg, Dagmar & Sabine Sczesny. 2001. Effekte des generischen Maskulinums und alternativer Sprachformen auf den gedanklichen Einbezug von Frauen. *Psychologische Rundschau* 52. 131–140. https://doi.org/10.1026//0033-3042.52.3.131.

Stahlberg, Dagmar, Sabine Sczesny & Friederike Braun. 2001. Name your favorite musician. *Journal of Language and Social Psychology* 20. 464–469. https://doi.org/10.1177/0261927X01020004004.

Stein, Simon David & Ingo Plag. 2021. Morpho-phonetic effects in speech production: modeling the acoustic duration of english derived words with linear discriminative learning. *Frontiers in Psychology* 12. https://doi.org/10.3389/fpsyg.2021.678712.

Tomaschek, Fabian, Peter Hendrix & R. Harald Baayen. 2018. Strategies for addressing collinearity in multivariate linguistic data. *Journal of Phonetics* 71. 249–267. https://doi.org/10.1016/j.wocn.2018.09.004.

Trömel-Plötz, Senta. 1982. *Frauensprache: Sprache der Veränderung*. Fischer Taschenbuch Verlag.

Trutkowski, Ewa & Helmut Weiß. 2023. Zeugen gesucht! Zur Geschichte des generischen Maskulinums im Deutschen. *Linguistische Berichte* 2023. https://doi.org/10.46771/9783967692792_2.

Völkening, Lena. 2022. Ist Gendern mit Glottisverschlusslaut ungrammatisch? Ein Analysevorschlag für das Suffix [ʔɪn] als phonologisches Wort. *Zeitschrift für Wortbildung / Journal of Word Formation* 6. 58–80. https://doi.org/10.3726/ZWJW.2022.01.02.

Widrow, Bernard & Marcian E Hoff. 1960. Adaptive switching circuits. In 96–104. IRE.

Wood, Simon N. 2017. *Generalized additive models: An introduction with R*. CRC Press. https://doi.org/10.1201/9781315370279.

Zacharski, Lisa & Evelyn C. Ferstl. 2023. Gendered representations of person referents activated by the nonbinary gender star in German: A word-picture matching task. *Discourse Processes* 60. 294–319. https://doi.org/10.1080/0163853X.2023.2199531.